THE EFFECTIVENESS OF MEDICAL SIMULATION IN TEACHING MEDICAL STUDENTS CRITICAL CARE MEDICINE: A PROTOCOL FOR A SYSTEMATIC REVIEW

Matt Beal, NORWICH MEDICAL SCHOOL
m.beal@uea.ac.uk
# CONTENTS

1.0 Background .......................................................................................................................... 3

2.0 Objectives ............................................................................................................................. 4

3.0 Criteria for Selecting Studies ............................................................................................... 4

4.0 Search Methods .................................................................................................................... 5

5.0 Methods of Review ............................................................................................................... 6

6.0 Ethical Considerations .......................................................................................................... 11

   6.1 Conflict of Interests ......................................................................................................... 12

7.0 Timetable .............................................................................................................................. 13

8.0 Discussion ............................................................................................................................. 14

9.0 Appendices ............................................................................................................................ 17

   9.1 Literature Search Plan .................................................................................................. 18

   9.2 Inclusion criteria ......................................................................................................... 19

   9.3 Data to be extracted ................................................................................................. 20

   9.4 Risk of Bias Assessment ......................................................................................... 21

6.0 References ........................................................................................................................... 14

9.0 Appendices ............................................................................................................................ 17
1.0 BACKGROUND

There is no common undergraduate curriculum in acute and emergency care (Shen et al., 2003) and deficiencies in knowledge are common amongst junior doctors (Smith and Poplett, 2002) who are often responsible for the early assessment and treatment of patients who are acutely ill (Smith et al., 2007). A review of undergraduate training in the care of acutely ill patients, found that training to be suboptimal and placing patients at risk (Smith et al., 2007).

At some point in medical education there is a need to hone skills on live patients; however this need must be carefully balanced against the ethical obligation to provide optimal treatment to patients (Ziv et al., 2003) and to maintain patient safety. In critical care, this ethical dilemma is further exacerbated as patients are often sedated or have reduced levels of consciousness, which may limit their ability to consent to being involved in this kind of education.

There is now a growing body of evidence relating to the use of Simulation Based Medical Education (SBME) (Smith et al., 2007), which may be of use in mitigating the ethical tensions that arise from using live patients as training tools for clinicians (Ziv et al., 2003).

Simulation is the process of recreating characteristics of the real world (Beaubien and Baker, 2004), which is useful because it allows the trainer to carefully control the learning environment, therefore optimising the situation for the skill being learnt.

The use of medical simulation to teach obstetric skills has been described, dating back to the 16th century (Buck, 1991). In the modern era, one of the earliest commercial mannequin simulators was the Resusci-Anne - developed in the 1960s by Asmund Laerdal (Cooper and Taqueti, 2008). The progression from these simple, low-fidelity mannequins has now led to the development of high-fidelity computerised mannequins, which are much more realistic.

Simulation has been shown to have a positive impact in a number of trainee groups (Davis et al., 2007, Hall et al., 2005, Wayne et al., 2008, Rosenthal et al., 2006), however its effectiveness during
the undergraduate medical years is not as clearly defined (Heitz et al., 2011). Reviews thus far have centred on post-graduate SBME, or in specialties other than critical care (Ilgen et al., 2013). The stage of professional development, as well as the skills being practiced may influence the effectiveness of the teaching method employed.

Exposure to SBME during medical school is highly variable and no studies have investigated an ideal amount of exposure time (Heitz et al., 2011). SBME is liked by medical students and faculty alike (Morgan et al., 2006, Morgan and Cleave-Hogg, 2000); however, the effectiveness of medical simulation vs. other teaching has been equivocal, with some papers reporting no difference, and others finding positive or negative effects (Gordon et al., 2006, Morgan et al., 2006), which is at odds with the literature from other domains and stages of professional development, which found moderate to large favourable effects (Ilgen et al., 2013).

Another review examining the effectiveness of virtual reality in surgery is currently being carried out (Aim and Ravaud, 2012). This review differs significantly from ours, concentrating on the use of screen-based simulators in learning surgical techniques, by a spectrum of professionals including medical students, surgical trainees and senior surgeons.

2.0 OBJECTIVES

The primary aim of this review is to determine the effectiveness of simulation-based training, in teaching medical students critical care medicine. The secondary aim of the review is to determine which type of simulation is most effective.

A systematic review of the research literature in the area of simulation-based medical education will be carried out.

3.0 CRITERIA FOR SELECTING STUDIES

For inclusion in the systematic review, studies should meet the following criteria:
**Participants** will be undergraduate medical students studying at a medical school recognised by the World Health Organization.

**Interventions** will be the use of simulation tools in critical care, intensive care, and resuscitation training; in adults, children or neonates. ‘Simulation Tools’ include Low-tech Simulators, High-tech Simulators, Screen-based Computer Simulators, Standardised Patients, Complex Task Trainers, Human Cadavers; Live Animals, Carcases or parts (See Appendix 8.2).

**Comparison** will be to either no-training, or other training modalities including but not exclusively: traditional lecture-based teaching, problem-based learning (PBL) and clinical shadowing. The use of any of these modalities as part of a simulation-based intervention will not necessitate exclusion, as most real-world applications would use a combined approach.

Studies comparing two groups with different forms of simulation, without a non-simulation control group, will also be included to allow direct comparison between simulation methods.

**Outcomes** will be any measure of knowledge or skill-based competency, in relation to managing critically unwell patients, including but not restricted to global rating scales, knowledge-based test scores and OSCE scores. Measures that are carried out both immediately or in the longer term will be included.

**Studies** will use randomised control trial (RCT) methodology. It is anticipated that there will be many studies of this design available, however if data from at least five controlled trails are not available, cohort studies will be included where there is a comparison group reported.

Other study methods will be excluded due to the potential for bias.

---

4.0 SEARCH METHODS
Six databases will be searched: ERIC, EMBASE, AMED, MEDLINE, The British Education Index and The Australian Education Index; from the year 1949 to current publications. Keyword and MESH term searches will be carried out (See Appendix 8.1). Keywords will be generated and checked by the author using an online medical thesaurus, and by checking keywords of pre-identified key papers to ensure maximum inclusion.

Bibliographies of included studies will be visually searched. Authors of included studies and manufactures of equipment for simulation will be contacted with a request for unpublished literature. Citation searching will also be conducted.

The search will be documented in detail at each stage to ensure replicability, and EndNote X4 will be used to record and manage the sources.

Towards the end of the writing up period, the literature search will be updated to ensure recent studies are not missed, and citation alerts will be set up on key papers.

5.0 METHODS OF REVIEW

Study Selection

Parallel independent assessment of paper eligibility will be carried out in two stages to minimize the risk of error (CRD, 2009). Stage 1: Two reviewers will screen titles and abstracts from the initial literature search to determine whether papers potentially meet the inclusion criteria, papers that clearly do not will be rejected. Stage 2: Two reviewers will review the full text of the remaining papers to determine eligibility in the review as per the inclusion criteria.

Disagreements at either stage will be resolved by discussion between the two reviewers. If the researchers fail to meet a consensus, a third reviewer will be consulted for arbitration.

The rate of reviewer agreement will be monitored, as evidence of poor agreement might suggest the need to revise selection criteria, or improve the coding process (CRD, 2009). Inter-assessor reliability
will be formally assessed using a Kappa statistic. Kappa is calculated by creating a contingency table of frequencies with each row and column indicating the categories of response for each observer. The frequencies that would be expected if the decisions were random, are calculated in a similar way to the chi-squared test, and the Kappa statistic is calculated comparing the proportion of agreement that would be expected by chance against the proportion that is observed (Petrie and Sabin, 2000).

The reasons for study rejection will be recorded for presentation in the final report.

**Quality Assessment**

Quality assessment to determine the risk of bias, will be carried out using the Cochrane Collaboration tool (CRD, 2009) *(See Appendix 8.4)*. The assessment will be published in the final report and studies determined to have a high level of risk will be excluded from the meta-analysis.

**Data Extraction**

Data extraction will be conducted by one reviewer, using Microsoft Excel (2007). A second reviewer will extract data from 30% of included studies to check accuracy. If there are inaccuracies identified, the second reviewer will continue to extract data from the rest of the eligible studies so that the entire dataset can be checked. Where disagreements occur, they will be resolved by discussion between the two reviewers. If the reviewers fail to meet a consensus, a third reviewer will be consulted for arbitration.

The data extraction sheet will be piloted before extraction starts to ensure it is easy to use and is consistent between reviewers.

Data will be extracted to ensure a complete record of the methodology, study design, participants, interventions, outcome measures and results. Maximal data extraction is planned to ensure that findings can be adequately followed up without returning to the original data set *(See Appendix 8.39.3 Data to be extracted)*.
Data Analysis

It is expected that the majority of studies identified will have used effect measures for continuous variables, most probably the mean difference, which is the absolute difference between the mean value in two groups (Higgins and Green, 2011). However, it is also anticipated that the studies may use different continuous scales. All data will therefore be calculated as Standardised Mean Differences (SMD) to allow for comparison. As the SMD does not correct for differences in the direction of the scale, corrections will be made to ensure that all scales point in the same direction.

The SMD expresses the size of effect in each study, relative to the observed variability regardless of the scales used:

\[
SMD = \frac{\text{Difference in mean outcome between groups}}{\text{Standard Deviation of outcome among participants}}
\]

The SMD may potentially be problematic in this review, because it makes the assumption that the Standard Deviations reflect differences in the measurement scale and not in the difference amongst the population, which may not be the case.

The inverse variance fixed-effect method for meta-analysis will be carried out, assuming no heterogeneity of the data. Heterogeneity refers to the variation of results between the studies and will be tested using the chi-squared statistic which assesses whether the observed differences are compatible with the expected differences through chance alone (Higgins and Green, 2011). If the data is found to be significantly heterogeneous (P=<0.10), careful attention will be paid to the source of variation (particularly in the delivery of the simulation) with a view to performing a random-effects meta-analysis, sub-group analysis, or conducting a narrative analysis where appropriate. It may be appropriate to exclude outlier studies if there is an obvious reason for the outlying result, however the potential for bias means that this should be avoided. Where potential sources of variation have been addressed or excluded, a p-value <0.10 will be sufficient evidence to indicate using the random effects model for meta-analysis.
The primary meta-analysis will compare the effectiveness of simulation against control. Six secondary analyses will be carried out to compare the effectiveness of each type of simulation intervention against the other simulation interventions: High fidelity simulators, Low fidelity simulators, Screen-based simulators and standardised patients. Where there are fewer than three studies for each comparison, a narrative synthesis will be carried out.

The immediate learning effect of simulation based teaching, as well as the influence on long-term retention is of interest in this review. Defining when the transition from short term memory to long term occurs is difficult and many different classifications exist (Tulving, 1995, Larsen, 2009, Roediger and Karpicke, 2006).

For the purposes of both meta-analyses, all studies will be pooled at the outcome closest to 24 hours post-simulation. This time point has been chosen as the most appropriate cut off for outcome measures aimed at assessing the immediate learning effect, and should therefore show the greatest and most-significant effect size. However, we appreciate that studies that only measure long-term knowledge retention may be a significant and problematic source of heterogeneity in this analysis, as skill retention declines over time (Smith et al., 2008).

For the primary meta-analysis, additional sub-group analysis will be carried out comparing time of outcome assessment post-simulation, type of simulation, type of control intervention, undergraduate year of student, and amount of simulation exposure. The categories for sub-group analysis will be:

**Time to outcome assessment** will pool studies with outcomes assessed at: up to 72 hours; between 72 hours and one month; and studies over one month post-simulation, using the latest time point presented in each study respective of these categories.

**Type of Simulation** will be categorised as high-fidelity, low fidelity, screen-based; human or animal cadaver; standardised patients, complex task trainers.
**Type of Control Intervention** will be categorised as no intervention, problem-based learning, lectures, clinical shadowing, self-directed study.

**Undergraduate level of student** will be categorised as junior (years 1-3) and senior (years 4 and above). Studies with students assessed from multiple levels will be pooled by the most senior students included.

**Amount of Simulation exposure** will be categorised as up to 4 hours; between 4 and 8 hours; and over 8 hours of simulation.

Data will be presented in a mixture of tables and forest plots generated using Cochrane’s RevMan:

- **Table 1** will present the characteristics of included studies: the methods, participants, intervention, outcomes and risk of bias.

- **Table 2** will present the pooled data and the primary meta-analysis: the number of participants in control and intervention arms, SMDs, Standard Deviations, Confidence Intervals, and P-Values for each arm of each study; the overall effect and heterogeneity statistic.

- **Table 3** will present a summary of the primary meta-analysis sub-grouped by type of simulation, including number of pooled participants, pooled SMD with confidence intervals and P-value.

- **Table 4** will present a summary of the primary meta-analysis sub-grouped by outcome measure, including number of pooled participants, pooled SMD with confidence intervals and P-value.

- **Table 5** will present the pooled data and the secondary meta-analyses: the number of participants in the different simulation arms, SMDs, Standard Deviations, Confidence
Intervals, and P-Values for each arm of each study; the overall effect and heterogeneity statistic for each simulation comparison.

**Figure 1** will be a study flow diagram, as outlined in the PRISMA statement.

*Figure 2* will be a forest plot of the main results comparing simulation against no simulation, including by study: number of participants, type of simulation, the raw data, point estimates and confidence intervals of SMD, p-Value as blocks, lines and text; heterogeneity statistics, and results of meta-analysis.

*Figure 3* will be a forest plot of the results sub-grouped by type of simulation, comparing simulation with no-simulation including by study: number of participants, type of simulation, the raw data, point estimates and confidence intervals of SMD, p-Value as blocks, lines and text; heterogeneity statistics, and results of meta-analysis for each sub-group.

*Figure 4* will be a forest plot of the secondary results comparing simulation against simulation, including by study: number of participants, the raw data, point estimates and confidence intervals of SMD, p-Value as blocks, lines and text; heterogeneity statistics, and results of meta-analysis for each simulation comparison.

### 6.0 Ethical Considerations

Systematic Reviews generally do not need ethical approval, however care will be taken to ensure that the review is ethically carried out. In particular, the methodology is based on accepted review techniques published in the Cochrane handbook (CRD, 2009), and conforms to the PRISMA guidelines (Moher et al., 2009). This is to ensure that the effect of bias is minimised in the review.

The main outcome measure of the review, effectiveness, is typically measured in terms of the overall improvement within the population. This is a consequentialist approach that takes no account of the distribution of individual or sub-group benefits within the population – so it is important to interpret
any results with a degree of caution, and to carry out sub-group analysis where appropriate. This ensures that the conclusions reflect what is actually demonstrated.

6.1 Conflict of Interests

The chief investigator would like to declare that he/she does not have any direct or indirect interests, financial or otherwise, that may conflict with the purposes of this review.
### Table 1: Simulation Review Timetable V1.1

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Start Date</th>
<th>Duration (Days)</th>
<th>Overrun (Days)</th>
<th>End Date</th>
<th>Resources</th>
<th>Time between tasks (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol Review</td>
<td>13/05/2013</td>
<td>14</td>
<td>15</td>
<td>11/06/2013</td>
<td>MB, CA, LH</td>
<td>1</td>
</tr>
<tr>
<td>Lit. Search</td>
<td>12/06/2013</td>
<td>3</td>
<td>0</td>
<td>15/06/2013</td>
<td>MB</td>
<td>1</td>
</tr>
<tr>
<td>AbTi Screening</td>
<td>16/06/2013</td>
<td>5</td>
<td>0</td>
<td>21/06/2013</td>
<td>MB, CA</td>
<td>1</td>
</tr>
<tr>
<td>Obtain FT</td>
<td>22/06/2013</td>
<td>10</td>
<td>0</td>
<td>02/07/2013</td>
<td>MB</td>
<td>1</td>
</tr>
<tr>
<td>Screen FT</td>
<td>03/07/2013</td>
<td>10</td>
<td>0</td>
<td>13/07/2013</td>
<td>MB, CA</td>
<td>1</td>
</tr>
<tr>
<td>Data Extract</td>
<td>14/07/2013</td>
<td>10</td>
<td>0</td>
<td>24/07/2013</td>
<td>MB, CA, TM</td>
<td>1</td>
</tr>
<tr>
<td>Data Synthesis</td>
<td>25/07/2013</td>
<td>10</td>
<td>0</td>
<td>04/08/2013</td>
<td>MB</td>
<td>1</td>
</tr>
<tr>
<td>Write-up</td>
<td>05/08/2013</td>
<td>30</td>
<td>0</td>
<td>04/09/2013</td>
<td>MB</td>
<td>1</td>
</tr>
<tr>
<td>Reviewer input</td>
<td>05/09/2013</td>
<td>20</td>
<td>0</td>
<td>25/09/2013</td>
<td>MB, CA</td>
<td>1</td>
</tr>
<tr>
<td>Final Edit</td>
<td>26/09/2013</td>
<td>3</td>
<td>0</td>
<td>29/09/2013</td>
<td>MB</td>
<td>1</td>
</tr>
<tr>
<td>Protocol Review</td>
<td>13/05/2013</td>
<td>14</td>
<td>15</td>
<td>11/06/2013</td>
<td>MB, CA, TM,LH</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1: Gantt chart for Medical Simulation Systematic Review
8.0 DISCUSSION

Variations in the way interventions were delivered between studies may cause quite significant heterogeneity of results. This may be problematic in the interpretation of those results, because we may find that the interventions turn out to be incomparable. This could be tackled by making the inclusion criteria more specific – however, given the variability in SBME in the real world, doing so would reduce the generalizability of the review.

There is also likely to be variation in vocabulary used to describe the same interventions, which means that screening of studies is likely to be difficult and close attention will need to be paid to reviewer agreement and potentially inclusion/exclusion criteria may need to evolve as new information is gained.

Variations between outcome measures used by the different studies – many of which will be unvalidated are likely to have a negative effect on the generalizability of the review because it will make it harder to identify the specific situations that the results actually apply. However, the variation could conversely be a good thing as the level of variation that we are likely to find between the different studies – probably reflects the variation that would be seen in the general population.

6.0 REFERENCES


9.0 APPENDICES
### TABLE 3: Literature Search Plan

<table>
<thead>
<tr>
<th>No.</th>
<th>Purpose</th>
<th>Search Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Population Keywords</strong></td>
<td>Medical Student$.tw OR Student Doctor$.tw OR Medical School.tw OR Undergraduate Doctor$.tw OR Undergraduate Medical Education.tw OR Medical Education.tw OR Training.tw OR Undergraduate.tw OR Student Physicians.tw</td>
</tr>
<tr>
<td>2</td>
<td><strong>Population MESH</strong></td>
<td>exp Medical Education/ OR exp Clinical Education/ OR exp Education/</td>
</tr>
<tr>
<td>3</td>
<td><strong>Intervention Keywords</strong></td>
<td>Simulat$.tw OR Manikin$.tw OR Moulage.tw OR Mannequin.tw</td>
</tr>
<tr>
<td>4</td>
<td><strong>Intervention MESH</strong></td>
<td>exp Simulation/ OR exp Computer Simulation/ OR exp Disease Simulation/ OR exp Simulator/ OR exp Simulation Model/</td>
</tr>
<tr>
<td>5</td>
<td><strong>Cochrane RCT filters</strong></td>
<td>randomized controlled trial.pt. OR controlled clinical trial.pt. OR randomized.ab. OR placebo.ab. OR clinical trials as topic.sh. OR randomly.ab. OR trial.ti.</td>
</tr>
<tr>
<td>6</td>
<td><strong>Cochrane RCT filters</strong></td>
<td>exp animals/ not humans.sh.</td>
</tr>
<tr>
<td>7</td>
<td><strong>Domain Area Keywords</strong></td>
<td>Critical Care.tw OR Cardiopulmonary Resuscitation.tw OR Intensive Care.tw OR Resuscitation.tw OR Emergency Treatment.tw</td>
</tr>
<tr>
<td>8</td>
<td><strong>Domain Area MESH</strong></td>
<td>exp Intensive Care/ OR exp Emergency Treatment/ OR exp Resuscitation/</td>
</tr>
<tr>
<td>9</td>
<td><strong>Population final</strong></td>
<td>1 or 2</td>
</tr>
<tr>
<td>10</td>
<td><strong>Intervention final</strong></td>
<td>3 or 4</td>
</tr>
<tr>
<td>11</td>
<td><strong>Study Design final</strong></td>
<td>5 not 6</td>
</tr>
<tr>
<td>12</td>
<td><strong>Domain Final</strong></td>
<td>7 or 8</td>
</tr>
<tr>
<td>12</td>
<td><strong>Final Search</strong></td>
<td>9 and 10 and 11 and 12</td>
</tr>
</tbody>
</table>
9.2 Inclusion criteria

**TABLE 4: Inclusion criteria**

<table>
<thead>
<tr>
<th>Included Interventions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-tech Simulators</td>
<td>Models or mannequins used to practice simple physical manoeuvres or procedures.</td>
</tr>
<tr>
<td>Simulated/Standardised patients</td>
<td>Actors trained to role-play patients, for training and assessment of history taking, physicals, and communication skills.</td>
</tr>
<tr>
<td>Screen-based computer simulators</td>
<td>Programs to train and assess clinical knowledge and decision making. <em>(E.g. Perioperative critical incident management, problem based learning, physical diagnosis in cardiology, acute cardiac life support.)</em></td>
</tr>
<tr>
<td>Complex Task Trainers</td>
<td>High-fidelity visual, audio, touch cues, and actual tools that are integrated with computers. Virtual reality devices and simulator that replicate a clinical setting. <em>(E.g. ultrasound, bronchoscopy, cardiology, laparoscopic surgery, arthroscopy, sigmoidoscopy, dentistry.)</em></td>
</tr>
<tr>
<td>Realistic patient simulators</td>
<td>Computer Driven, full-length mannequins. Simulated anatomy and physiology that allow handling of complex and high-risk clinical situations in lifelike settings, including team training and integration of multiple simulation devices.</td>
</tr>
<tr>
<td>Cadavers</td>
<td>Cadavers or Cadaveric material used to practice physical manoeuvres or procedures.</td>
</tr>
<tr>
<td>Animal Simulators</td>
<td>Live Animals, Carcases or parts used to practice physical manoeuvres or procedures.</td>
</tr>
</tbody>
</table>

*Taken and adapted from (Ziv et al., 2003)*

*Note: This list may not be exhaustive.*
9.3 Data to be extracted

Source
- Study ID (created by review author).
- Review author ID
- Citation and contact details.
- Year of publication
- Journal Name
- Title

Eligibility
- Confirm eligibility for review.
- Reason for exclusion.

Methods
- Study design.
- Total study duration.
- Sequence generation*.
- Allocation sequence concealment*.
- Blinding*.
- Other concerns about bias*.

Participants
- Total number.
- Setting.
- Stage of profession
- Age.
- Sex.
- Country.
- [Date of study].

Interventions
- Total number of intervention groups.

For each intervention and comparison group of interest:
- Specific intervention.
- Intervention details (sufficient for replication, if feasible).
- Integrity of intervention

Outcomes
- Outcomes and time points (i) collected; (ii) reported*.

For each outcome of interest:
- Outcome definition
- Validated Measure?
- Unit of measurement
- For scales: upper and lower limits, and whether high or low score is good.

Results
- Number of participants allocated to each intervention group.

For each outcome of interest:
- Sample size.
- Missing participants*.
- Summary data for each intervention group (e.g. 2×2 table for dichotomous data; means and SDs for continuous data).
- Estimate of effect with confidence interval; P value
- Subgroup analyses

Miscellaneous
- Funding source.
- Key conclusions of the study authors.
- Miscellaneous comments from the study authors.
- References to other relevant studies.
- Correspondence required.
- Miscellaneous comments by the review authors.

Taken and adapted from: (CRD, 2009)
## 9.4 Risk of Bias Assessment

<table>
<thead>
<tr>
<th>Source of Bias</th>
<th>Low Risk</th>
<th>High Risk</th>
<th>Unclear Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>12.1 Random Sequence Generation</strong></td>
<td>Describes a random component in the sequence generation process. Such as:</td>
<td>Describes a non-random component in the sequence generation process. A systematic, non-random approach, for example:</td>
<td>Insufficient information about the sequence generation process to permit judgement of ‘Low risk’ or ‘High risk’</td>
</tr>
<tr>
<td></td>
<td>- Referring to a random number table;</td>
<td>- Sequence generated by odd or even date of birth;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Using a computer random number generator;</td>
<td>- Rule based on date (or day) of admission; or on hospital or clinic record number.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Coin tossing;</td>
<td>By judgement or some method of non-random categorization of participants, for example:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Shuffling cards or envelopes;</td>
<td>- Allocation by judgement of the clinician; or preference of the participant;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Throwing dice;</td>
<td>- Allocation based on the results of a laboratory test or a series of tests;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Drawing of lots;</td>
<td>- Allocation by availability of the intervention.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Minimization.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>12.2 Allocation Concealment</strong></td>
<td>Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:</td>
<td>Participants or investigators enrolling participants could possibly foresee assignments, such as allocation based on:</td>
<td>Insufficient information to permit judgement of ‘Low risk’ or ‘High risk’. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed</td>
</tr>
<tr>
<td></td>
<td>- Central allocation (including telephone, web-based and pharmacy-controlled randomization);</td>
<td>- Using an open random allocation schedule (e.g. a list of random numbers);</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Sequentially numbered drug containers of identical appearance;</td>
<td>- Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered);</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Sequentially numbered, opaque, sealed envelopes</td>
<td>- Alternation or rotation; Date of birth; Case record number;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Any other explicitly unconcealed procedure.</td>
<td></td>
</tr>
<tr>
<td><strong>12.3 Blinding of Participants and Personnel</strong></td>
<td>Any one of the following:</td>
<td>Any one of the following:</td>
<td>Any one of the following:</td>
</tr>
<tr>
<td></td>
<td>- No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding;</td>
<td>- No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding;</td>
<td>- Insufficient information to permit judgement of ‘Low risk’ or ‘High risk’;</td>
</tr>
<tr>
<td></td>
<td>- Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken</td>
<td>- Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.</td>
<td>- The study did not address this outcome.</td>
</tr>
<tr>
<td><strong>12.4 Blinding of outcome assessment</strong></td>
<td>Any one of the following:</td>
<td>Any one of the following:</td>
<td>Any one of the following:</td>
</tr>
<tr>
<td></td>
<td>- No blinding of outcome assessment, but the review authors judge that the outcome measurement is not</td>
<td>- No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding;</td>
<td>- Insufficient information to permit judgement of ‘Low risk’ or ‘High risk’;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- The study did not address this outcome.</td>
</tr>
</tbody>
</table>
likely to be influenced by lack of blinding;  
- Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.

12.5 Incomplete Outcome Data

<table>
<thead>
<tr>
<th>Any one of the following:</th>
<th>Any one of the following:</th>
<th>Any one of the following:</th>
</tr>
</thead>
</table>
| - No missing outcome data;  
  - Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias);  
  - Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups;  
  - For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate;  
  - For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size;  
  - Missing data have been imputed using appropriate methods. | - Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups;  
  - For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate;  
  - For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size;  
  - ‘As-treated’ analysis done with substantial departure of the intervention received from that assigned at randomization;  
  - Potentially inappropriate application of simple imputation. | - Insufficient reporting of attrition/exclusions to permit judgement of ‘Low risk’ or ‘High risk’ (e.g. number randomized not stated, no reasons for missing data provided);  
  - The study did not address this outcome. |

12.6 Selective Reporting

<table>
<thead>
<tr>
<th>Any of the following:</th>
<th>Any of the following:</th>
<th>Insufficient information to permit judgement of ‘Low risk’ or ‘High risk’. It is likely that the majority of studies will fall into this category</th>
</tr>
</thead>
</table>
| - The study protocol is available and all of the study’s pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way;  
  - The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon). | - Not all of the study’s pre-specified primary outcomes have been reported;  
  - One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified;  
  - One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);  
  - One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis;  
  - The study report fails to include results for a key outcome that would be expected to have been reported for such a study. | |
| 12.7 Other Bias | The study appears to be free of other sources of bias | There is at least one important risk of bias. For example, the study:  
- Had a potential source of bias related to the specific study design used; or  
- Has been claimed to have been fraudulent; or  
- Had some other problem | There may be a risk of bias, but there is either:  
- Insufficient information to assess whether an important risk of bias exists; or  
- Insufficient rationale or evidence that an identified problem will introduce bias |

(Based on Cochrane tool, (CRD, 2009))